THE EFFECTS RETROTRANSPOSON INSERTION HAS ON THE METHYLATION DYNAMICS OF THE INTEGRATION SITE

by

Clement O. Ojo Jr

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
April 2020

Abstract: Non-LTR retrotransposons are the only known class of active mobile genetic elements in the human genome. These elements are endogenous mutagens that can contribute to disease pathologies through a variety of mechanisms. One of the least documented mechanisms, by which active retrotransposons can contribute to disease, is through the creation of new CpG island (CGIs)s at their insertion sites. In fact, more than half of all CGIs in the human genome are thought to be derived from retrotransposons. There is evidence that suggests that repetitive DNA elements played a critical role in the establishment of genome-wide methylation patterns and that DNA methylation of CpGs evolved primarily as a host defense mechanism against transposable elements. Despite this fact, not much is known about the epigenetic impact that retrotransposons can have on their insertion sites. To gain a better understanding of this impact, we looked at the effect that non-LTR retrotransposons have on the methylation status of flanking DNA in three different systems. The first system involved an L1 insertion into the 5' UTR of the androgen receptor gene of a patient that caused Partial Androgen Insensitivity Syndrome. The second system involved an antisense SVA insertion into intron 32 of the TATA-binding protein-associated factor-1 (TAF1) gene that caused X-linked dystonia parkinsonism. The third system involved full-length L1 insertions introduced into a human ovarian teratocarcinoma cell-line. In all 3 systems, we hypothesized that the presence of the new non-LTR retrotransposon insertion would alter the methylation status of the flanking DNA CpG sites both upstream and downstream of the insertion sites. We have obtained preliminary data that has laid the groundwork for the successful completion of this research at a later date following the reopening of the university.

Primary Readers and Advisor: Dr. John Goodier, Dr. Haig Kazazian, Jr.
Secondary Reader: Dr. Anna Coppola

Acknowledgements:

I want to first thank my family for their continued support of me in all endeavors in which I choose to undertake. I would like to thank Dr. Haig Kazazian and Dr. John Goodier for their steady mentorship and guidance throughout my undergraduate and graduate study of retrotransposons. It was a pleasure to work with them and I benefited greatly from their wisdom, critiques and advice. I would like to acknowledge Alisha Soares for her technical support in many aspects of this project. I would also like to thank Katsumi Yamaguchi, a fellow researcher in this lab, for his many timely suggestions. To all individuals who graciously provided samples for use in this study, I would like to thank them. Lastly, I would like to acknowledge everyone involved in the masters' program this year for their role in this process.

Table of Contents

List of Tables:

List of Figures:

Chapter 1: A review: the role of retrotransposon insertion in disease

It has been roughly seventy years since Barbara McClintock first discovered the existence of transposable elements (TEs)in maize [1]. Since then, there has been a lot of research and tools developed to better understand the biological significance of mobile DNA in the genome. Formerly considered "junk DNA," it is now understood that these elements have played a huge role in genome evolution and that they can contribute to disease [11-13]. This review will focus on a portion of what is currently understood about the biology of transposons and their role in human disease.

TEs, also known as "jumping genes", are DNA sequences that can mobilize from one genomic location to the next. The ability of TEs to "jump" is mediated by key endogenous proteins such as DNA transposase and reverse transcriptase. TEs that mobilize using gene-encoded proteins are known as autonomous elements. TEs that rely on autonomous element proteins to mediate their mobilization in *trans* are known as non-autonomous elements. Altogether, these transposons constitute approximately half of the mammalian genome. At least two-thirds of the human genome is thought to be repetitive or repeat-derived [2,7,8].

TEs are divided into two groups based on their mechanisms of genomic insertion. The first group are the DNA transposons; these mobilize via a DNA intermediate. The second group are the retrotransposons; these mobilize via an RNA intermediate. See Figure 1.

I. DNA transposons

DNA transposons have relatively low copy numbers and occupy only 3% of the human genome [3]. These elements are inactive in all primates; the only known exception was found to be a family of piggyBac elements in bats [4]. There are several families of DNA transposons present within mammalian genomes. These families are classified based on their transposase and the biochemical mechanism of transposition that they employ [5]. Generally, all DNA

transposons mobilize via a “cut and paste” mechanism. This process involves the self-excision of the DNA element from its current site, followed by its integration at a new target site. There are a couple of essential steps, in the biochemistry of this process that all known families of DNA transposons share [14,15]. At the donor site, the exposure of 3’ hydroxyl groups at the transposon’s ends always occurs. At the target site, a strand transfer reaction, driven by a nucleophilic attack of the exposed 3’ hydroxyl group also always occurs. This results in the integration of the element at its target site [5]. The target site tends to be close to the original donor locus - a phenomenon known as “local hopping” [16].

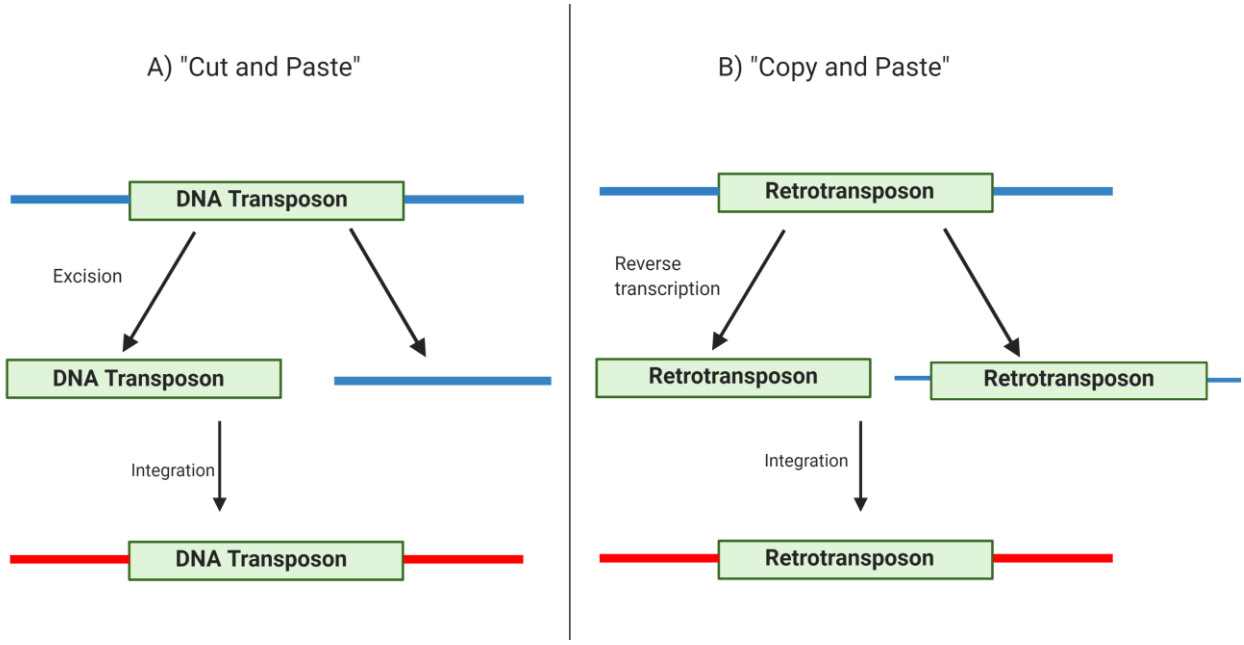


Figure 1.1: Mechanisms of Transposition [31].

II. Retrotransposons

These elements mobilize within the genome via a “copy and paste” mechanism (Fig. 1). The process involves the transcription of the retrotransposon into RNA, the reverse transcription

of the RNA into cDNA and the integration of the cDNA copy into a new genomic site [6]. Historically, retrotransposons are subdivided into two groups based on whether they possess long-terminal repeats (LTR) or not (non-LTR).

LTR retrotransposons

LTR elements have been found in the genomes of many eukaryotic organisms (e.g. Ty1 in yeast) [19,20]. They make up ~8% of the human genome and are believed to be mostly inactive; HERV-K may be the exception [6]. LTR elements are very similar to retroviruses, which suggests they share a common evolutionary history [21]. Both have been found to possess the *gag* and *pol* genes, which encode the proteins necessary for their replication. The main structural difference is that LTR elements, unlike retroviruses, lack a fully functional *env* gene. This means that they can only mobilize within their genome of origin [18]. In humans, LTR elements are found to exist in truncated and mutated forms or as solo LTRs [5]. This is the reason why LTR elements are generally considered to be inactive in humans, although the existence of active elements has not been ruled out [45,46].

There are some common components in the life cycle of a typical LTR retrotransposon. The cycle begins with the transcription of the element by RNA Polymerase II at an internal promoter in the 5' LTR region. The RNA is then subsequently transported to the cytoplasm where it is translated. The resulting proteins form a virus-like-particle (VLP) containing the RNA. The reverse transcriptase protein then reverse transcribes the retrotransposon RNA into a cDNA copy. This complete cDNA is transferred back into the nucleus where integration takes place [21]. See Figure 2.

Non-LTR retrotransposons

Non-LTR retrotransposons are the only class of TEs that are still active in the human genome [13, 23]. The long interspersed element-1 (LINE-1 or L1) is the only autonomous active

non-LTR element. The other non-LTR elements (e.g. short interspersed element (SINE) Alus and SVAs) are non-autonomous; they cannot mobilize without L1 proteins.

L1s

As the only autonomous non-LTR elements, L1s occupy approximately 17% of the human genome. Most L1s are inactive due to mutations, rearrangements and truncation events. However, on average, every human has roughly 80-100 L1 elements that are still active [23]. An active L1 is ~6 kb in length and has a ~900 bp 5' untranslated region (UTR), a ~200 bp 3' UTR, 2 open reading frames (ORF1 and ORF2) on the sense strand, and a 3' UTR. The 5' UTR of an L1 acts as an internal promoter. The L1 3' UTR has a polyA tail and a weak polyA signal (AATAAA) which is responsible for transcription termination [6, 23]. ORF1 encodes a ~40 kDa RNA-binding protein with chaperone activity, that binds RNA molecules as a trimer [47]. ORF2 encodes a ~150 kDA protein with endonuclease (EN) and reverse transcriptase (RT) activities. Both proteins, ORF1p and ORF2p, are essential for L1 retrotransposition. L1s also have an antisense promoter in the 5' UTR that encodes a short ORF0p product of unknown function [24].

The L1 retrotransposition cycle begins in the nucleus with the transcription of a bicistronic mRNA. There are several transcription factors involved in this process. Following transcription, L1 mRNA is translated in the cytoplasm to produce ORF1p and ORF2p [25]. These proteins assemble with L1 RNAs to form a ribonucleoprotein particle (RNP). Some of these RNPs then re-enter the nucleus, where an integration mechanism termed target-primed reverse-transcription (TPRT) takes place [26]. The genomic site of integration is largely determined by the endonuclease (EN) activity of ORF2p. This enzyme recognizes a consensus sequence, of the form 5'-TTTT/AA-3' and cleaves the bottom strand of the target site to free up a 3' hydroxyl [23]. The L1 RNA polyA tail anneals to the single overhang produced and ORF2p uses the free 3' hydroxyl as a primer for reverse transcription. This generates the first strand of

L1 cDNA. Second-strand cleavage and synthesis occur soon afterwards, and the structure is resolved and integrated into the genome by cellular mechanisms that have not been clearly defined. Upon TPRT completion, the retrotransposon insertion is characterized by target site duplications (TSD), a hallmark of this process. Occasionally during TPRT, a process called "twin priming" occurs, which helps to explain L1 inversion events [27]. See Figure 2.

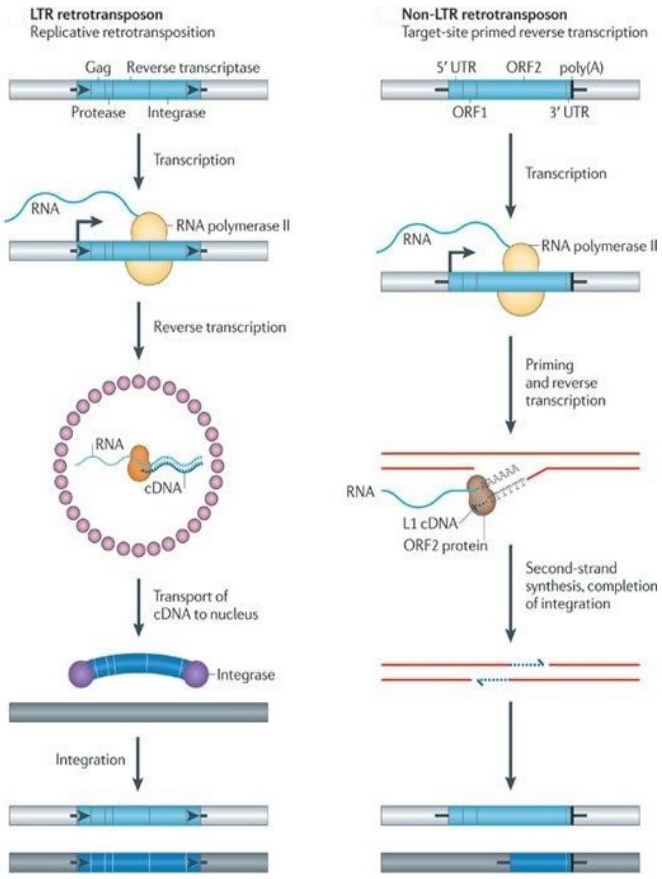


Figure 1.2: The Mechanisms of Retrotransposon Integration [64]

Alus

Alus constitute ~10% of the human genome. This noncoding short-interspersed element (SINE) is ~300 bp long. It is also the most abundant retroelement in the human genome with every human possessing over 1 million copies on average [32]. Structurally, Alus are dimeric with left and right monomers connected by an A-rich linker. The left monomer contains an

internal RNA polymerase III promoter, with A and B boxes, for transcription [23]. Alus are dependent on L1 proteins for mobilization. Cell retrotransposition assays have shown that Alus require L1 ORF2p to mobilize in *trans* [28]. While L1 ORF1p has been found to not be essential, it does, however, enhance Alu retrotransposition [33]. Alus, like L1s, upon insertion are characterized by poly A tails of varying length and TSDs. Both of these are hallmarks of TEs that undergo the TPRT process. See Figure 2.

SVAs

The other non-autonomous non-LTR retroelement in humans is the SINE-VNTR-ALU (SVA). This hominid-specific element comprises approximately 2% of the human genome, and is the youngest active retrotransposon [6]. The structure of an SVA, 5' to 3', consists of: a) hexameric repeats $(CCCTCT)_n$, b) an antisense Alu-like domain, c) a GC-rich variable number of tandem repeats (VNTR) domain, and d) a SINE-R domain derived from an extinct HERV-K [28]. Only L1 ORF2p is required for SVA retrotransposition. However, L1 ORF1p has, in retrotransposition assays, been found to enhance SVA insertions [29,48]. SVA genomic insertions also contain hallmarks of L1-mediated retrotransposition, including both the presence of a polyA tail of varying length and TSDs (See Figure 3).

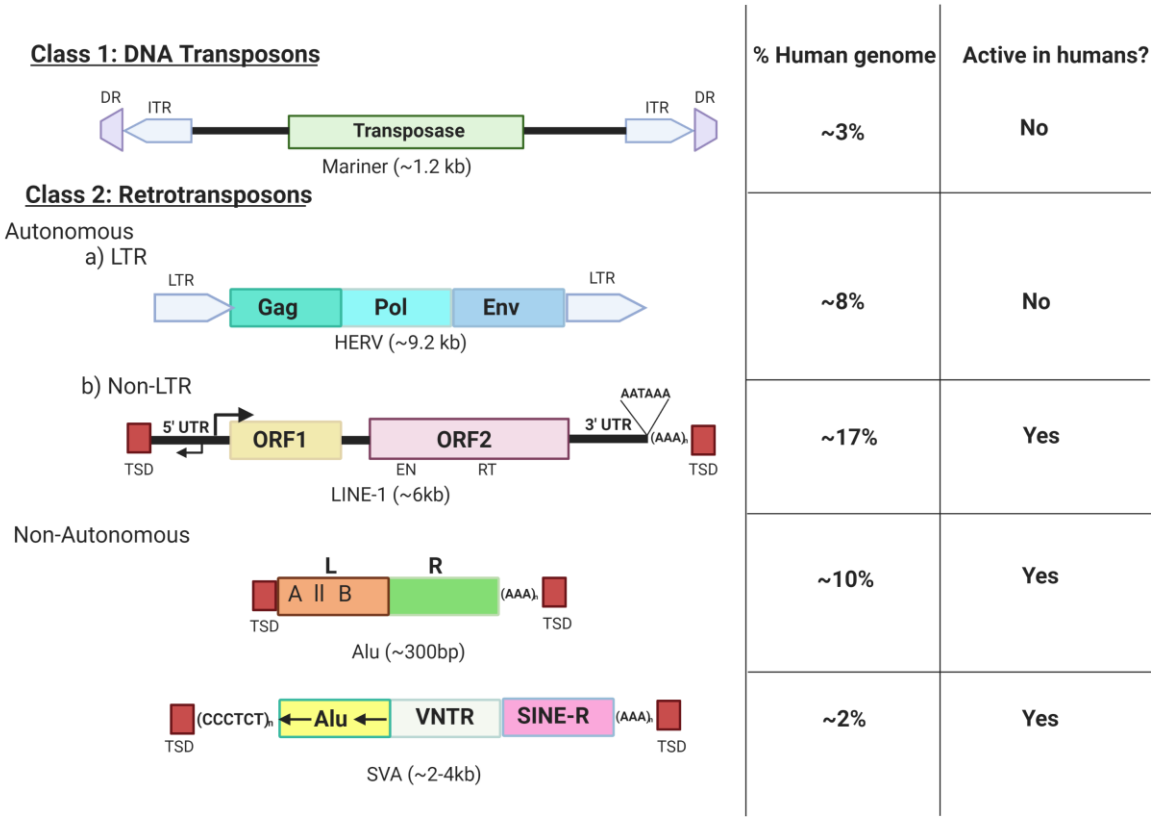


Figure 1.3: The Different Types of Transposable Elements in Mammals. [Adapted from (30,31)]. Abbreviations: *Mariner*: DR, direct repeat; ITR, inverted terminal repeat; *HERV*: LTR, long terminal repeat; Gag, group-specific antigen; Pol, polymerase; Env, envelope; *LINE-1*: RT, reverse transcriptase domain; EN, endonuclease domain; TSD, target site duplication; *Alu*: L, Left monomer; R, Right monomer; *SVA*: VNTR, variable number tandem repeats, SINE-R, short interspersed element of HERV origin.

III. The role of Retrotransposons in human disease

In 1988, Dr. Haig Kazazian and colleagues provided the first piece of evidence that human disease could be caused by *de novo* retrotransposon insertions [34]. In one patient case, they found a truncated L1 insertion in a boy with hemophilia A. The insertion was present within exon 14 of the factor VIII gene; a gene located on Chromosome X. This same insertion was not found in the parents' genome who were normal. However, the mother did have what was determined to be the progenitor full-length insertion present on her chromosome 22 [30,34]. This evidence suggested that L1 insertion events could occur spontaneously in the gametes or during early embryogenesis. It suggested that L1s are still active in the human genome today and that they could cause disease. Now we know that every human has roughly 80-100 potentially active

L1 elements in their genome [23]. Moreover, any two humans differ on average by ~300 L1 insertions [11].

There are now >130 documented cases of retrotransposition events associated with human disease [30]. These cases include both de novo and inherited occurrences, and germline and somatic insertion events. Each documented case has contributed to our understanding of retrotransposon biology and the mechanisms by which insertions can cause/contribute to disease.

Mechanisms by which Insertions can cause/contribute to disease

There are a variety of mechanisms by which retrotransposons can cause or contribute to disease [11]. The mechanism observed most often is insertional mutagenesis. This occurs when a retrotransposon inserts into a gene or regulatory region and directly disrupts its normal functioning. This leads to readily observable phenotypes. Historically, this mechanism is exemplified by the first reported case of a disease-causing insertion in the germline (the hemophilia A case) [34]. An additional historical example of insertional mutagenesis would be the first documented instance of a somatic insertion causing disease [35]. This case involved L1 insertion into the last exon of the APC gene, a tumor suppressor gene, causing a sporadic case of colorectal cancer.

In addition to disrupting genomic sequences upon insertion, retrotransposons can contribute to genomic instability by causing deletions and by transducing flanking genomic sequences upon insertion [49-51]. Transduction events occur in both SVAs and L1s to varying degrees. In 3' transduction, the RNA transcription machinery bypasses the weak polyA signal present in its 3'UTR and uses an alternative polyA signal downstream in the 3' flanking sequence. This means L1s have the potential to mediate exon shuffling (if they mobilize an

exon/regulatory sequence with them as they insert) [37]. In 5' transduction, a promoter upstream of the retrotransposon sequence transcribes through the retrotransposon.

Alus and L1s insertions can also contribute to disease through ectopic recombination (ER). ER events can lead to several types of genomic rearrangements (e.g. deletions, duplication, translocations, etc.). For example, a nonallelic homologous recombination event, between two antisense Alu elements, caused a ~5 kb deletion in a region of the Low-density lipoprotein receptor gene of a patient, which resulted in hypercholesterolemia [39].

There are many mechanisms in which a TE can contribute to genomic instability and cause disease. Therefore, it makes sense that host organisms have multiple defense mechanisms to prevent their spread.

Natural Defense Mechanisms against Retrotransposon Insertions

Organisms have developed several defense mechanisms to prevent retrotransposons from generating genomic instability and consequently disease. These mechanisms are enacted at various stages of development and have been found to involve many restriction factors (e.g. APOBECs, MOV10, etc.) [6]. A main mechanism of retrotransposon restriction is gene silencing. Gene silencing occurs within primordial germ cells (PGCs). DNA methyltransferases and associated proteins (e.g. Dnmt3L, DNMT3A) act in concert to methylate the 5' UTR of retrotransposons, which significantly suppresses TE action. Within the germline, retrotransposon activity is regulated at several levels by the small interfering RNA PIWI/piRNA pathway [6,23]. At the chromatin level, the PIWI/piRNA pathway promotes genome-wide DNA and histone methylation of TEs, which greatly suppresses their activity. At the post-transcriptional level, PIWI/piRNA complex can cleave TE transcripts via RNA interference. There are several other

natural defense mechanisms that organisms employ to restrict RI events. There is a constant battle between TEs and their host organisms.

IV. Area of study: DNA Methylation and non-LTR retrotransposons

We hypothesized that *de novo* non-LTR retrotransposon insertions could affect the methylation status of flanking DNA. There is evidence to support this thinking. Both L1s and SVAs are known to possess CpG islands (CGIs) near their 5' termini [52]. See Figure 1,4. Of the 28 million CpG dinucleotides present in the human genome, 27% and 12% are contained in SINES and LINEs, respectively [53]. SVAs are ~ 60% GC-rich, exceeding 70% GC content within their VNTR region [54]. Additionally, more than 60% of human SVAs are present in or within 10 kb of genes [55]. Thus, each new full-length SVA or L1 insertion introduces novel CGIs within the vicinity of genes. CGIs are potential sites for methylation and gene silencing by enzymes.

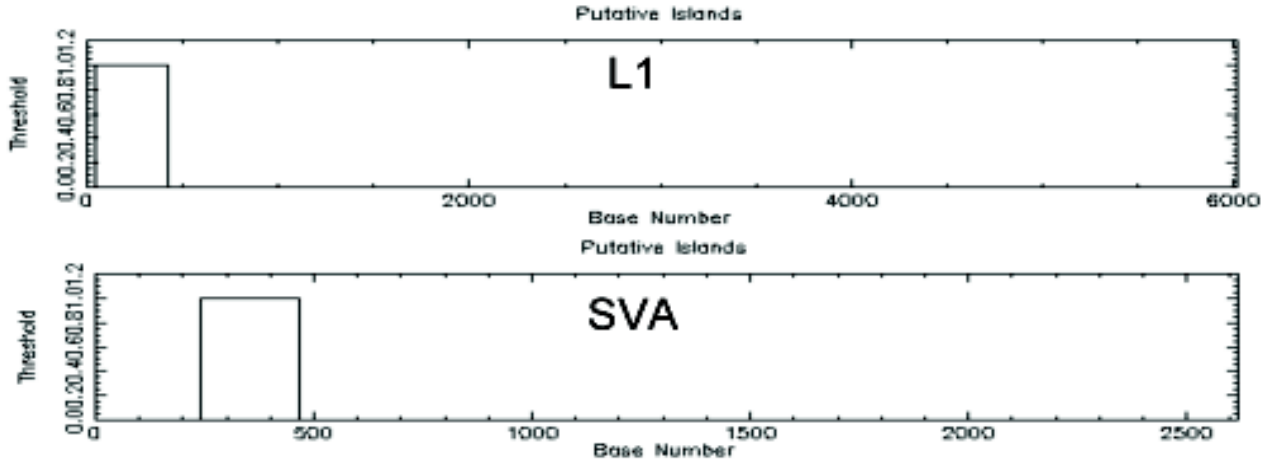


Fig. 1.4. CpG islands predicted in L1 and SVA by the CPGplot algorithm. The SVA CpG island (CGI) is present within its Alu-like region. The L1 CGI is present in its 5' UTR region.

There is evidence, especially from plants, that suggests that repetitive elements play a critical role in the establishment of genome-wide methylation patterns. For example, methylation

spreading *in cis* from plant retrotransposons into flanking DNA regions has been found [58, 59]. The occurrence of this phenomenon, in mammals, has not been studied as extensively. However, it was found that mouse B1 SINEs can act as "methylation centers" [60]. Additionally, methylation spreading from human Alus into flanking DNA has been implicated in the silencing of some tumor suppressor genes [61]. There is also evidence that SVAs can affect the methylation status of CpGs present in flanking DNA regions. In one study, a 60% loss of methylation in DNA upstream of a human polymorphic SVA element was observed [62]. There have also been cases where L1s insertions have been observed to affect the methylation status of flanking CpG residues [63].

To test the hypothesis that new non-LTR retrotransposon insertions can affect the methylation status of neighboring DNA, we used whole genome bisulfite conversion, site-specific PCR, TOPO-TA cloning, and Sanger sequencing to examine methylation effects in three different systems involving unique full-length non-LTR retrotransposon insertions:

1) an L1 insertion into the 5' UTR of the androgen receptor gene of a patient that caused Partial Androgen Insensitivity Syndrome (Batista et al 2019). See Chapter 2.
2) an SVA insertion in antisense orientation into intron 32 of the TATA-binding protein-associated factor-1 (TAF1) gene that caused X-linked dystonia parkinsonism, a rare movement disorder that almost exclusively afflicts males of Filipino descent (Makino et al. 2007). TAF1 gene encodes the largest subunit of the TFIID transcription complex. See Chapter 3: future study #1.
3) 5 clonally derived human ovarian teratocarcinoma PA1 cell lines, each containing a full-length L1 introduced into their genomes by a cell culture retrotransposition protocol. See Chapter 3: future study #2.

Chapter 2: Methylation Analysis of DNA flanking L1 in PAIS case

Introduction

Androgen Insensitivity Syndrome (AIS) is the most common cause of Disorders of Sexual Development (DSD) in individuals with a 46, XY karyotype. It is usually caused by mutations in the androgen receptor (AR) gene. This gene, located at Xq11-12, has 8 exons that encode the androgen receptor [41]. It is responsible for regulating the expression of genes involved in sexual development and differentiation. In AIS, AR mutations result in varying degrees of undervirilization of the external genitalia. AIS clinically presents itself in 3 forms: complete (CAIS), partial (PAIS) and mild (MAIS). In CAIS, a 46, XY individual has completely female external genitalia. In MAIS, a 46, XY individual has completely male external genitalia associated with infertility and possible gynecomastia [41]. The PAIS clinical phenotype is a large spectrum, of undervirilization, between the two extremes of CAIS and MAIS. Out of all 3 forms of AIS, PAIS has been found to be the most difficult to diagnose [43]. Yet, it is also the most common form of AIS with an estimated incidence rate of 1 in 130,000 births [40]. One reason why PAIS is difficult to diagnose is because causative AR gene mutations, in coding regions and splice sites, have only been identified in <25% of patients who have a clinical diagnosis of PAIS. This greatly contrasts with CAIS, in which the mutation has been identified within the gene in >90% of cases [41]. This suggests that factors outside of the AR gene, which influence its' expression, may contribute to PAIS. In the literature, this has been referred to as AIS type II [42]. By looking at mutation-negative cases of AIS, we can gain a better insight into its etiologies.

Long-interspersed elements (L1s) belong to a class of DNA elements in the human genome called transposable elements (TE). TEs are mobile elements that can insert into new locations in the genome. L1s are the only active autonomous retroelement and is responsible for mobilizing

other elements such as SVAs and Alus. Retrotransposon insertion (RI) has been found to cause or contribute to several somatic and germline diseases. One of the main mechanisms, by which retrotransposons can cause disease is insertional mutagenesis. Insertional mutagenesis typically involves the disruption of a gene via insertion into a coding region, a splice site or some other regulatory region [6]. This is the predominant mechanism observed in disease-causing RI cases. Another potential mechanism, by which active RIs can cause disease, is by causing epigenetic changes at the integration site. There is supporting evidence for this mechanism. For example, in teratocarcinoma cell lines, alterations in local histone modifications have been observed following L1 insertion suggesting that L1s could potentially epigenetically impact a target site [43]. By exploring cases in which RI may affect the epigenetic profile of an insertion site, we gain more knowledge about the biological properties of retrotransposons and the mechanisms by which they could cause disease.

A former researcher in our lab, Dr. Rafael Batista, discovered a case of PAIS where whole exon sequencing of the AR gene showed no mutation. This fact seemingly conflicted with the observation that AR gene expression, in this patient, was significantly impaired [41]. Upon further analysis, a L1 insertion was found in the 5' UTR of the patient's AR gene. Familial analyses revealed that the mother was a carrier for the insertion, however, was unaffected (Table2.1). We hypothesized that the significant reduction in AR gene expression observed in this patient was likely due to the epigenetic effect the L1 insertion has on the flanking DNA at the insertion site. We expected to see a significant difference in the methylation status of CpGs flanking the 5' UTR of the PAIS patient's full-length L1 insertion than seen at the same CpGs in a comparable control lacking the insertion (the patient's normal XY brother).

Material and Methods:

AIS Genetic Analysis Studies: We obtained genomic DNA from the PAIS patient from Rafael Batista at the University of São Paulo. The DNA was originally obtained from peripheral blood leukocytes using the proteinase K-SDS salting-out method. The entirety of the Androgen Receptor (AR) gene was sequenced using primers designed via the Primer3 bioinformatics program [75]. The PCR products were prepped following the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit protocol (Life Technologies Corp., Carlsbad, CA). The products were analyzed using the ABI Prism Genetic Analyzer 3130XL (LifeTechnologiesCorp.).

Bisulfite Primer Design: Several bisulfite primers were designed with the aid of MethPrimer bioinformatics tool and Primer3Plus [75,76]. These primers were between 20-30 bp in length and were designed to preferably amplify regions between 200-500 bp. The first set of primers were designed so that the 5' AR flank and L1 would be amplified. A second set of primers were designed so that the 3' AR flank and the 3' end of the L1 would be amplified. A third set of primers were designed to amplify the empty site AR DNA for cases without the L1 insertion, including the promoter core region. (Fig 2.2). All these primers were designed such that they would amplify their respective regions with at least 3 CpG sites present within an amplicon. Due to limited amounts of patient DNA, we first tested the third set of primers in RPE-1 DNA. The RPE-1 DNA modeled the no insertion case. We amplified the "empty site" in the RPE-1 DNA. We also amplified the region just downstream of where the L1 insertion would be if the RPE cells had the insertion. This region is where the AR gene promoter core in the 5'UTR is located.

Bisulfite DNA sequence analysis: Bisulfite conversions of the DNAs were performed using the EpiTect Fast bisulfite kit (Qiagen) according to the manufacturer's instructions. To determine the DNA methylation status, this treatment converts unmethylated cytosine resides to thymine, while

those that are methylated (5-methylcytosines) are resistant to treatment and stay as cytosines. Normally, mostly all non-CpG cytosines are not methylated in the human genome. This allows the calculation of the conversion efficiency of the bisulfite process. We calculate the conversion efficiency by obtaining the ratio of converted to non-converted non-CpG Cs in each sample. A DNA sample with all non-CpG cytosines sites converted to thymines are considered to have a conversion efficiency of 100%.

Bisulfite-Nested PCR:

Bisulfite-treated DNAs were amplified by a nested PCR protocol. Reagents from the Epitect MSP kit (QIAGEN) were used. PCR was performed in a volume of 10 µl containing: 2x EpiTect Master Mix, 5 µM of each primer and 10 ng of BS-treated DNA. Products of expected size were run on a 1% agarose gel.

1st BS PCR conditions:

1) 95 °C for 10 min
2) 94 °C– 15 sec
3) 50-58 °C – 30 sec
4) 72°C – 2 min

x 30

5) 72°C– 10 min

Used 1st PCR product as a template for second PCR.

2nd BS PCR conditions:

1) 95 °C for 10 min
2) 94 °C– 15 sec
3) 55 °C –30 sec
4) 72°C –2 min

x 30

5) 72°C– 10 min

DNA Purification & Sequencing: PCR products of the expected size were extracted and purified following the QIAquick gel extraction kit (Qiagen) protocols. Purified DNAs were cloned into a vector following the TOPO TA Cloning kit protocol (pCR2.1, Invitrogen) and then transformed into NEB 5-alpha competent cells on agar plates containing 40 ul of X-gal. These plates were kept overnight at 37°C. After ~24 hours in the incubator, white colonies were picked in order to perform colony PCR.

Colony PCR Parameters:

1) 95 °C – 2:30 min
2) 95 °C – 30 sec
3) 58°C – 30 sec
4) 72°C – 1 min

x 30

5) 72°C – 3 min

The Qiagen Miniprep Kit protocol was followed in order to extract and purify the DNA for Sanger sequencing. All sequencing was done by the JHMI Synthesis & Sequencing Facility.

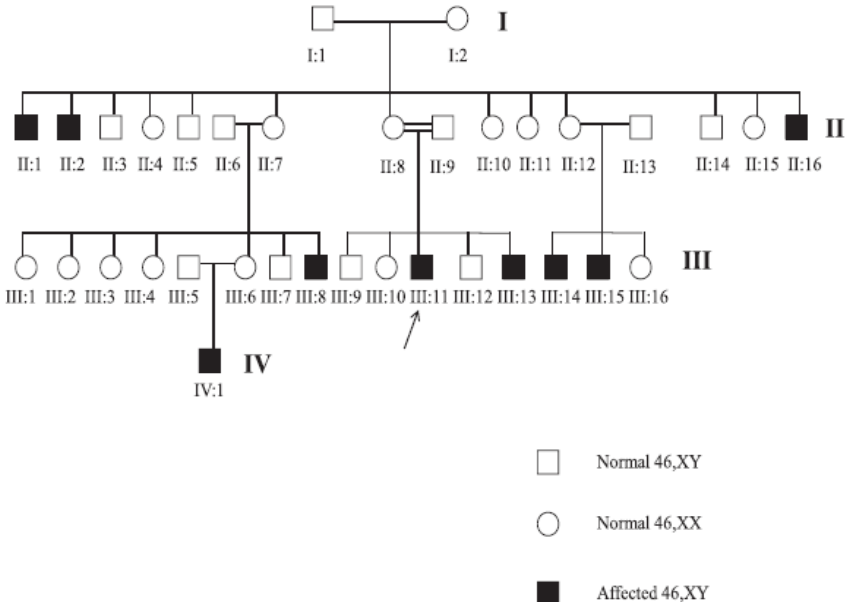


Figure 2.1: AIS Family Pedigree. Arrow points to the PAIS patient (with the L1 insertion) [41].

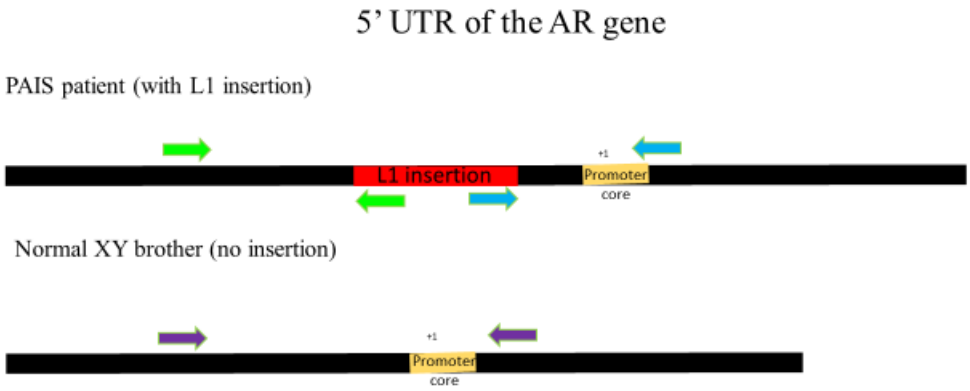


Figure 2.2: PAIS primer set schematic. The first set of primers were designed to amplify the 5' AR gene flank and the 5' L1 end (green arrows). The second set of primers were designed to amplify the 3'AR flank and the 3' end of the L1 (blue arrows). The third primer set was designed to amplify regions in the 5' UTR in cases with no insertion; the empty site + promoter core (purple arrows).

Primer Set Name	Primer Name	5' to 3'	Product size (bp)	Total CpGs
Empty site	5' AR UTR (out-F)	ATTTTAGTGGATATTGAATTTGGAAG	290	
	5' AR UTR (out-R)	CCCTTCTCTTACTCAAAAAAATTCAAC		
	5' AR UTR (in-F)	TTAGTGGATATTGAATTTGGAAGG	286	8
	5' AR UTR (in-R)	CCCTTCTCTTACTCAAAAAAATTCA		
5' AR promoter region -no insertion	5' AR. Promoter (out-F)	GTTGAATTTTTTTGAGTAAGAGAAGGG	410	
	5' AR Promoter (out-R)	CCTCATCCAAAACCAAATAACC		
	5' AR Promoter. (in-F)	TGAATTTTTTTGAGTAAGAGAAGGG	370	14
	5' AR Pro. (in-R)	GTGAGGATGGTTTTTTTTAAGTTTA		
5' AR UTR region - L1 insertion	5' AR UTR L1 insert. (out-F)	GGGAGTTAGTTTGTTGGGAGAG	275	
	5' AR UTR L1 insert. (out-R)	CAATCCTACCAAACACTTTCCTTAC		
	5' AR UTR L1 insert. (in-F)	GGAGTAAGTTTAGAGGTAGAGGAG	239	18
	5' AR UTR L1 insert. (in-R)	TCCTACCAAACACTTTCCTTACTTCC		

Table 2.1: RPE DNA Primer set combinations.

Results:

To test whether the presence of the L1 insertion affected the methylation status of flanking DNA in the 5' UTR of the AR gene, we planned to perform site-specific bisulfite PCR in the PAIS patient and his normal XY brother. We designed 3 primers sets. The first primer set was designed to amplify the 5' end of the L1 and the 5' flanking DNA in the PAIS patient. The secondary primer set was designed to amplify the 3' end of the L1 and the 3' flanking DNA in the PAIS patient. The third set of primers was designed to amplify the empty site of the 5' UTR in cases with no insertion (e.g. normal XY brother). We planned to compare the methylation status of CpGs in the flanking DNA for the case with the L1 insertion (PAIS) to the case without the insertion (normal XY brother). We wanted to see if there were any significant differences in the methylation levels of the CpGs in DNA flanking the insertion.

Due to limited amounts of patient DNA available for testing, we first performed preliminary experiments in genomic DNA extracted from RPE-1 cells (Retinal epithelial cells immortalized with hTERT). The RPE-1 DNA did not have an L1 insertion present in the AR gene, so it served as a model for the no insertion case. In the RPE-1 DNA, we successfully amplified the empty site and assessed the methylation status of the CpG sites (Fig. 2.3). The overall conversion efficiency was found to be ~99.3%. We also amplified the AR gene core promoter region present in the 5' UTR in the RPE cells (Fig 2.4). We assessed the methylation status of the CpG sites present in this region. The overall conversion efficiency was found to be ~99.7%.

We were able to amplify a DNA region upstream of the L1 insertion in the 5' UTR of the PAIS patient. We assessed the methylation status of the CpG sites (Fig 2.5). The overall conversion efficiency was found to be ~100%.

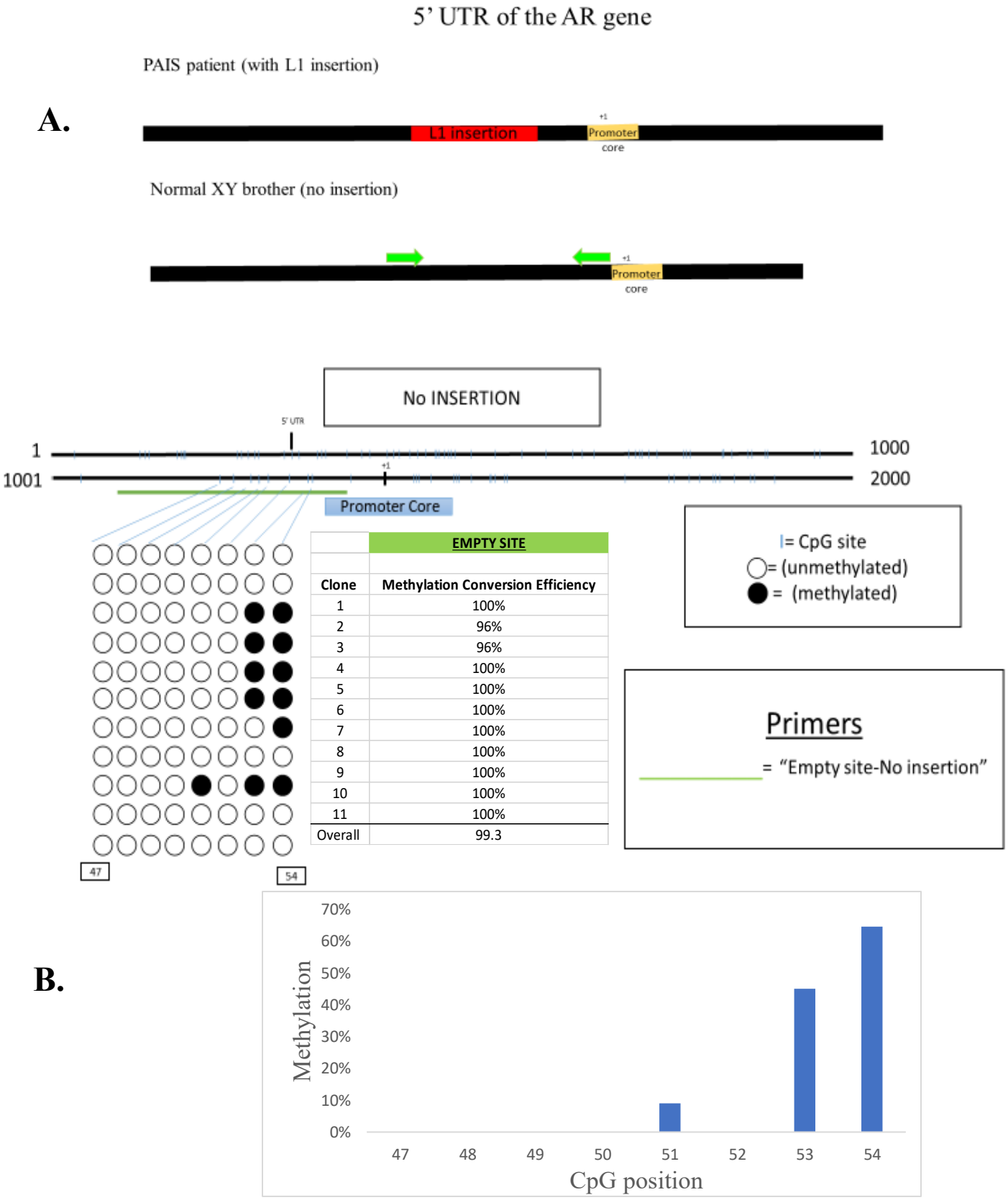
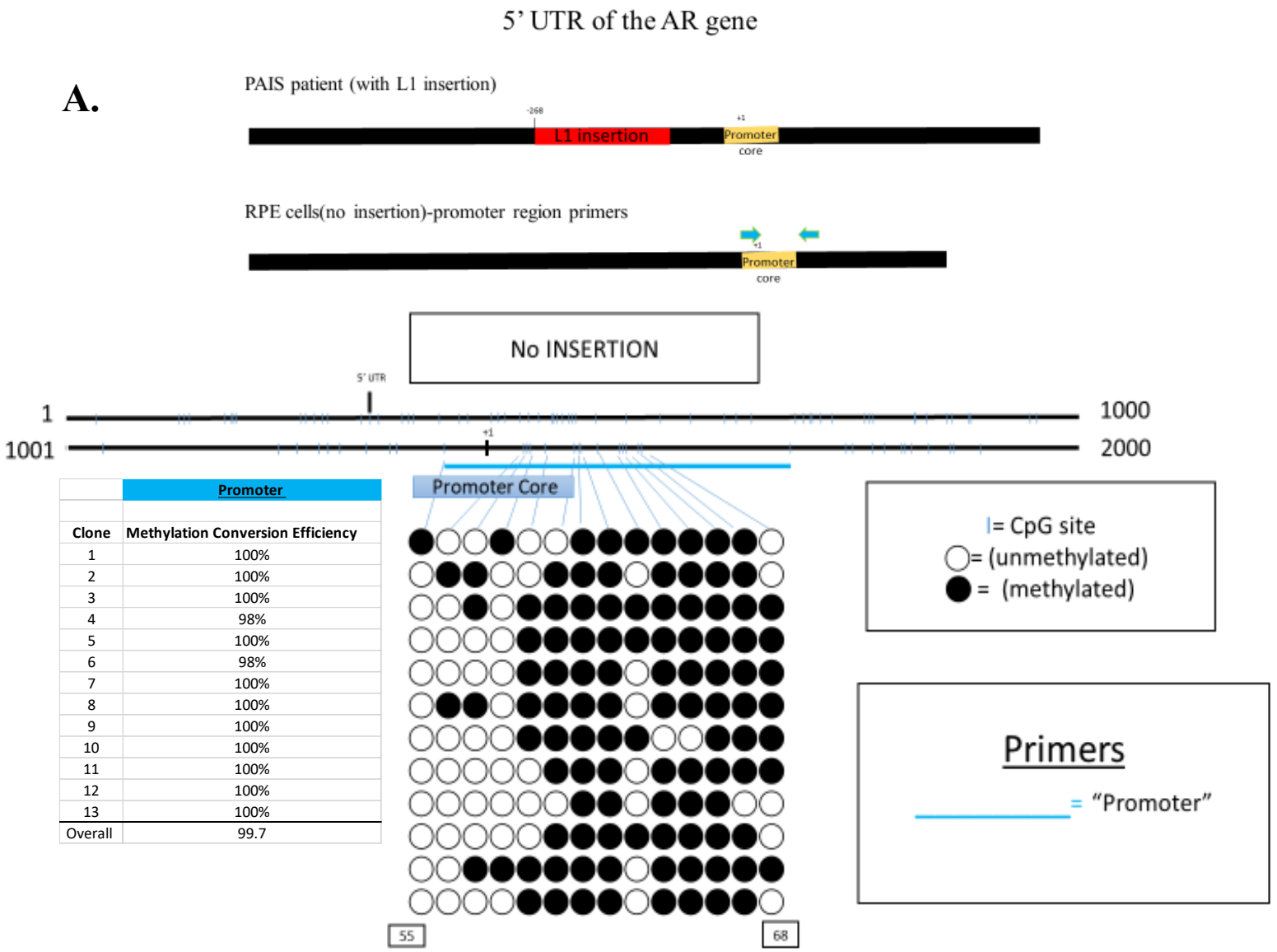


Figure 2.3: RPE, No insertion schematic- empty site. A) The "empty site," the region flanking where the L1 insertion would be present in the PAIS case, was amplified in the RPE DNA. The overall conversion efficiency was ~99.3%. B) The data on the methylation status at each CpG site, in the 5' UTR, of the AR gene was obtained.



	Promoter
Clone	Methylation Conversion Efficiency
1	100%
2	100%
3	100%
4	98%
5	100%
6	98%
7	100%
8	100%
9	100%
10	100%
11	100%
12	100%
13	100%
Overall	99.7

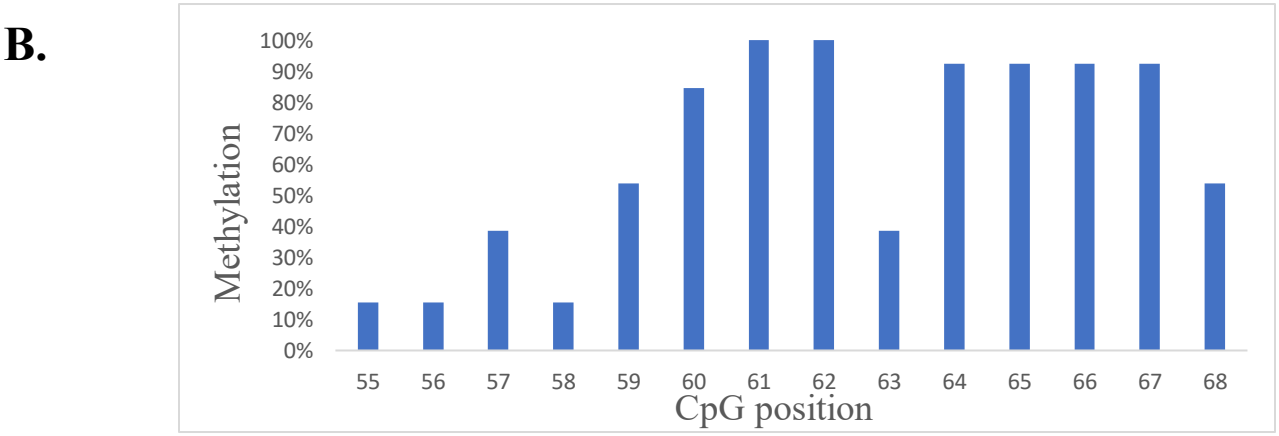


Figure 2.4: RPE, No insertion case- core promoter region. A) The "promoter" region was amplified in the RPE DNA. This region is adjacent to where the L1 insertion would be in the PAIS case. The overall conversion efficiency was ~99.7%. B) The data on the methylation status at each CpG site, in the 5' UTR, of the AR gene was obtained.

A.

cgagatttcggggagttagtttgttgggagagcgggacggttcggagtaagtttagaggtagaggaggcgatagagggaaaaaggg
tcgagttagtcgttttagtgttgtataggagtcgaagggacgtattacgttagttttagttcggttttagcgatagttaacgtttt
ttgtagcgcggcggtttcgaagtcgtcgttcggagttgttttttttttttcggtgaagttttaaaagttgttaaagattcggaggg
aagtaaggaaagtgtttggtaggattgacggttgttttgttttttttttttttattcgttttttttattttgttttttttttt
ttttcgttttttttttcgtagttgttttagtcggttatttttagttaatttttttattatttttttttttattcgttttttcgt
tttcgtcggtttagcgttgttagttcgagtttgtagagaggtaatttttttggttgcgagcgggcgagttagttgtatattgtaa
agaaggttttaggagttaggcgattggggagcggttttagtattgtagttacgattcgtttggttaggttgtacgcggagagaat
tttttgttttttttattttttttttattttttttgttttttttatttcgagtgcggagttagagattaaaagatgaaaaggtag
ttaggttttagtagttaaaaaataaaataaataaaaataaaaaagtcgaaataaaagaaaaagataataatttagtttttatttg
tatttattttagtggatattgaatttggaaggtggaggatttgttttttttttttAAGATTTGGGTATTTTTTTTTTTTTTTT
AGATATTTTTTAAAAGAAGATATTTATGTAGTTAAAAAATATATGAAGAAATGTTTATTATTATTGGTTATTAGAGAAATGTAAAT
TAAAATTATTATGAGATATTATTTTATATCGGTTAGAATGGTAATTATTAAGAAGTTAGGAAATAATAGGTGTTGGAGAGGATGCG
GAGAAATAGGAATATTTTTATATTGTTGGTGGGATTGTAAATTAGTTTAATTATTGTGGAAGTTAGTGTGGCGATTTTTTAGGGAT
TTAGAATTAGAAATATTATTTGATTTAGTTATTTTATTATTGGGTATATATTTAAATGAGTATAAATTATGTTGTTATAAAGATAT
ATGTATACGTATGTTTATTGCGGTATTATTTATAATAGTAAAGATTTGGAATTAATTTAAATGTTTAATAATGATAGATTGGATTA
AGAAAATGTGGTATATATATATTATGGAATATTATGTAGTTATAAAAAATGATGAGTTTATATTTTTTGTAGGGATATGGATGAAA
TTGGAAATTATTATTTTTAGTAAATTATCGTAAGAATAAAAAATTAAATATCGTATATTTTTATTTATAGGTGGGAATTGAATAAT
GAGATTATATGGATATAGGAAGGGGAATATTATATTTTGGGGATTGTGGTGGGGTCGGGGGAGGGGGGNAGGGATAGTATTGGGAG
ATATATTTAATGTTAGATGATATATTAGTGGGTGTAGCGTATAAGTATGGTATATGTATATATATGTAATTAATTTGTATAATGTG
TATACGTATTTTAAAATTTAGAGTATAATAAAAAAAAAAAAAAAAAAAAANNNNNNNAAAAAAAAAAAAAAAAAAAAGATTTGGGTAT
TTTTTgaatttatttttaagtattaagagatagattgtgagtttagtagggtagattttgtttatcgtgtgttttttttgtacg
agattttgaggttgttagagcgttttttgcgtggttgttttcgtaagttttttttttggagttttcgtaggtgggtagttagtt
gtagcgattatcgtattattatagttgttgaatttttttgagtaagagaaggggaggcgggtaagggaagtaggtggaagattt
agttaagtttaaggatggaagtgtagttagggttgggaagggtttatttcggtcgtcgtttaagatttatcgaggagtttttag
aatttgttttagagcggtgcgcgaagtgatttagaattcggggttttaggtatttagaggtcgcgagcgtagtatttttcggcgttag
tttgttgttgttgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtaagaga
ttagttttaggtagtagtagtagtagtagtagggtgaggatggttttttttaagtttatcgtagaggttttataggttatttggttttg
gatgaggaatagtaatttttatagtcgtagtcggttttggagtgttatttcgagagaggttgcgttttagagtttggagtcgtcgt
ggtcgttagtaaggggttgtcgtagtagttgttagtattttcggacgaggatgatttagttgttttatttacgttgtttttgttgg
gtt

5' UTR of the AR gene

B.

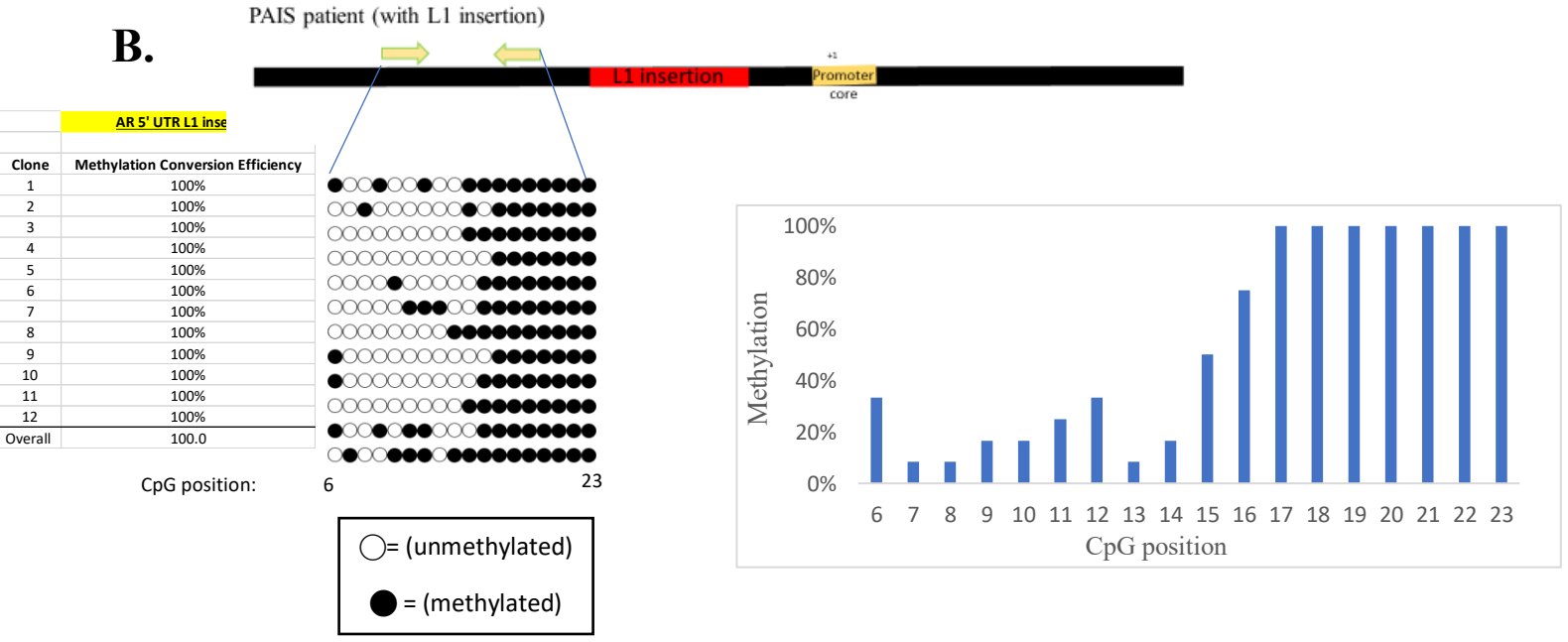


Figure 2.5: AR 5' UTR - L1 insertion case.

A) In the PAIS patient, a region of their 5' UTR was amplified. This region was upstream of where the L1 insertion was present. This region encompassed 18 CpG sites. The primer which denotes this site is highlighted in yellow).

Key: lowercase is the flanking genomic DNA, uppercase is the L1 sequence, highlighted in gray are the target site duplications (TSDs), in red are the CG sites, promoter core region in black.

B) The methylation status of the CpGs were assessed. Generally, CpGs closer to the insertion site were more frequently methylated.

Discussion:

One of the main challenges of bisulfite treatment of DNA is that it is an inherently harsh process. Treatment of DNA with bisulfite (HSO_3^-) results in the deamination of unmethylated cytosines (C) to uracil (U). However, during this process fragmentation of the DNA into small single stranded sequences, usually less than 500 nucleotides (nt) long, occurs [78]. For this study, we used the Epitect Fast Bisulfite Kit for the conversion process and the Epitect MSP kit for the site-specific bisulfite PCR. A study comparing the effectiveness of the various bisulfite conversion kits found that the EpiTect Kit recovers ~10.2% of the input DNA with average recovery fragment lengths of ~414 nucleotides [80]. Therefore, in order to maximize the probability of successfully amplifying a region of interest, the Epitect MSP kit recommends a high input concentration of DNA and amplicon sizes of up to ~250 bp. However, due to limited amounts of the patient DNA, we were forced to work with lower DNA concentrations. Additionally, the lengths of the amplicons that we were interested in generating were typically larger than 250 bp. Although more challenging, amplifying larger amplicons using lower concentrations of DNA is possible. Both the amplicon length and starting DNA concentration were factors that contributed to the difficulty we experienced in completing this project. This reality is the main reason why we first decided to model the no-insertion case, normal XY brother, in RPE-1 gDNA. By modeling this case in the RPE-1 DNA, we were able to successfully obtain a list of several primers that would have likely been able to amplify the regions we were interested in amplifying in the normal XY brother.

In this experiment we gained preliminary data on the methylation status of the CpGs upstream of the L1 insertion in the actual PAIS case. We also modeled what the no insertion case might look like in the XY normal brother by using RPE-1 gDNA. Initially, we hypothesized that the significant reduction in AR gene expression observed in the AR patient

was likely due to the epigenetic effect the L1 insertion has on the flanking DNA at the insertion site. Therefore, the results that we would expect to see would be a significant increase in methylation of CpGs present in DNA flanking the insertion in the PAIS patient as compared to the methylation of those same CpG sites in the normal XY brother. However, we were unable to obtain the data necessary for computing meaningful statistical analysis. When we were about to amplify the regions of interest in the actual normal XY brother, and obtain sequencing data, the emergence of COVID-19 halted our progress. Johns Hopkins mandated, due to the coronavirus pandemic, that all non-essential lab work be stopped, including access to sequencing facilities and reagents.

The successful completion of this project would have required methylation data from flanking CpG sites both upstream and downstream of where the L1 insertion is in both the normal XY male and PAIS patient. This data must include information about the methylation status of the same flanking CpG sites in the case with the insertion (PAIS patient) and the case without it (normal XY male). With the entirety of this data we would have been able to perform bisulfite conversion methylation analysis by using the Quantification tool for Methylation Analysis (QUMA). QUMA is a commonly used program that uses Fisher's Exact Test statistics to determine whether any differences in methylation levels are statistically significant [79]. QUMA would have allowed us to determine whether the presence of the L1 insertion (in the PAIS patient) led to statistically significant changes in the average methylation levels observed in each flanking CpG site (both upstream and downstream from the insertion).

In terms of results, we were expecting the average methylation levels of the flanking CpG sites to be significantly elevated in the presence of the L1 insertion (PAIS patient) relative to the absence of it (normal XY individual). We expected for this to be the case for CpG sites located

both upstream and downstream of the insertion. This result would have supported our hypothesis that the L1 insertion elevated the methylation level of the flanking DNA in the PAIS patient. This result would have also served as a possible explanation for why the PAIS patient had a significantly reduced AR gene expression profile without a mutation in a gene-coding/intronic region of their AR gene. This result would have also served as further evidence that an L1 can epigenetically affect its site of insertion and in some cases, contribute to disease phenotypes.

Another potential case study is the mother of the PAIS patient. The mother is a carrier for the L1 insertion in one of her X-chromosomes. This means that we could have theoretically looked at the methylation status of flanking CpG sites in the presence of the insertion and the absence of it (empty site) within the same individual. However, this comparative study would have been complicated by the phenomenon known as X-inactivation which occurs during embryonic development. This process involves the global inactivation of most genes present either on the maternally or paternally derived X-chromosome in each cell through methylation. New studies have come out that have demonstrated that skewed X-inactivation, where imbalanced inactivation of the X-chromosomes occurs, is common within the female population [81]. This phenomenon of X-inactivation, which is not completely understood, would make it difficult to attribute any significant differences in methylation status seen between insertion site and empty site to the presence of the L1 alone.

Chapter 3: Future Studies

Future Study #1: SVA in X-linked Dystonia Parkinsonism

Introduction: X-linked Dystonia Parkinsonism (XDP) is an adult-onset neurodegenerative movement disorder that almost exclusively affects males of Filipino descent [65,66]. The clinical phenotype usually presents itself in early adulthood as focal dystonia, which over time, becomes generalized to various regions of the body. As the disease progresses, the dystonia either coexists with or is replaced by parkinsonism [67]. This disease has no known cure. Most XDP patients are found to inherit an identical haplotype; variants of this haplotype are rare [70]. This haplotype consists of five specific nucleotide substitutions, a 48bp deletion and a full length antisense F-subtype SVA insertion within intron 32 of their TAF1 gene (Fig 3.1). The TAF1 gene encodes the TATA-binding protein-associated factor-1, which is the largest subunit of the transcription factor II D (TFIID) complex which is involved in transcriptional regulation [68, 69]. Even though most XDP patients share a common haplotype, there is a large amount of heterogeneity in the presentation and the progression of the disease (e.g. age of onset, disease severity and degree of cognitive dysfunction) [71-73]. The heterogeneity has been linked to the presence of the SVA insertion [72, 73].

SVAs are hominid-specific short interspersed nuclear elements (SINEs) that belong to the youngest family of active human retrotransposons (Fig. 3.2). Their name is an acronym reflecting their composite nature: human endogenous retrovirus (HERV)-K(HML-2)-derived SINE-R, variable-number-of-GC-tandem-repeats (VNTR), Alu-like region, and CCCTCTn hexamer array (Fig. 3.2). SVAs are ~ 60% GC-rich, exceeding 70% GC content within their VNTR region [54]. SVAs introduce new CpG islands (CGIa) where they insert. They have also been found to affect the methylation status of CpGs present in DNA regions flanking the insertion [62].

Based on all these known properties of SVAs, we hypothesized that the presence of the SVA insertion in intron 32 of the TAF1 gene increases the DNA methylation status of TAF1 DNA flanking the insertion sites of SVA-positive XDP patients. This phenomenon could explain the decrease in TAF1 gene expression experienced by individuals with XDP. If true then we would expect to see a significant difference in the methylation status of either: 1) X-chromosome DNA flanking the TAF1 SVA insertion of male XDP patients vs normal male controls (e.g. unaffected related individuals lacking the SVA insertion), or 2) the empty versus filled chromosomal sequences at the site of the SVA insertion of XDP carrier females.

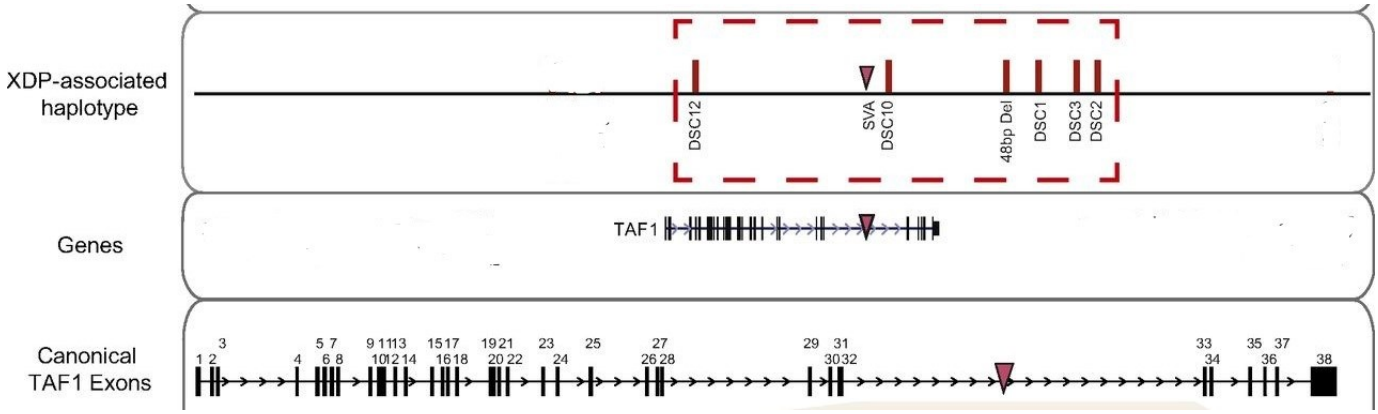


Figure 3.1: XDP haplotype. The haplotype is seen in the red dashed box. It consists of 5 single nucleotide substitution annotated as "DSC," a 48-bp deletion (48 bp Del), and an SVA insertion in intron 32 of the TAF 1 gene [Adapted from (41)].

Method: Several bisulfite primers were designed with the aid of MethPrimer bioinformatics tool and Primer3Plus [76,77]. These primers were between 20-30 bp in length and were designed to preferably amplify regions between 200-500 bp. The first set of primers were designed to amplify the region upstream of the SVA insertion in the XDP patients (this region contains 6 CpGs). The second set of primers were designed to amplify the region downstream of the SVA insertion in the XDP patients (this region contains 5 CpGs). The outer primer combinations

would be used to amplify the empty site. See Figure 3.2. This would allow us to compare the methylation status in these flanking regions in SVA-positive XDP patients to the same DNA regions in normal controls (SVA-negative relative).

Figure 3.2: XDP SVA schematic. The red arrows refer to primer set 1, to amplify regions upstream of the SVA insertion. The green arrows refer to primer set 2, to amplify regions downstream of the insertion. Use the outer forward primer and outer reverse primer to amplify the empty site.

Future Directions Discussion:

This project was very difficult mainly due to the highly repetitive nature of the SVA element and the limitations of the bisulfite conversion protocol used. The SVA is characterized by 5' hexamer arrays of (CCCTCT)n that vary in length in between patients (usually between 35 and 52 repeats) [44]. This fact alone makes primer design very challenging and it is often difficult for a PCR reaction to read through the entire repeat. In order to design unique primers on the 5' end, that amplify both the TAF1 flank and 5' SVA, it requires an amplicon to be a minimum of ~300 bp long in the best-case scenario (e.g. patient with 35 hexamer repeats, with a minimum of 3 CpG sites present within the flanking DNA). This length is an issue mainly due to the limitations of the bisulfite conversion procedure. In this experiment, we used the EpiTect MSP Kit for site-specific bisulfite PCR. The ideal amplicon size that they suggest for bisulfite converted DNA is ~250 bp [80]. During the conversion process, all unmethylated cytosines are converted into uracils. DNA also fragments into single stranded DNA due to the inherent harsh nature of this treatment. The fragmented DNA is usually around 500 nucleotides long and the DNA recovery rate varies [78]. For the kit that we used, a comparative study found that it recovers ~10.2% of the input DNA with average recovery fragment lengths of ~414 nucleotides

[80]. Therefore, the success of this experiment is largely dependent on how exactly the genomic DNA fragments during the conversion process.

Future Study #2: L1 in human ovarian teratocarcinoma cell lines

We obtained 5 clonally derived human ovarian teratocarcinoma PA1 cell lines, each containing a full-length L1 introduced into their genomes by a cell culture retrotransposition protocol [75]. These were obtained from our colleague, Dr. Jose Luis Garcia-Perez of the University of Edinburgh. In three of the five cell lines, the L1 was introduced into an intergenic region (e.g. RAM27, RAM272 and RAM 286). In the other two cell lines, an L1 was introduced into the intron of a gene (e.g. RAM29 and RAM 221). See Table 3.1.

Methods: We designed one set of bisulfite primers to amplify the upstream DNA flanking the L1 insertions. We designed a second set of bisulfite primers to amplify the downstream DNA flanking the L1 insertions (Fig 3.2). A third primer set was designed to amplify the empty site. This will allow us to compare the methylation status in these flanking DNA regions in chromosomes with the L1 insertion and those without the insertion (empty site). A significant difference in methylation status of the L1-flanking CpG sites would indicate that the presence of the L1 likely caused this difference.

CELL LINE	SUBLINE	ENZYME	TSD LENGTH	pA LENGTH	CHR LOCATION	TRUNCATION	OBSERVATIONS	LENGtH	GENE	ORIENTATION	EN cleavage
PA-1	RAM27	SSP1	12	A54	Xp22.2	1	intergenic	8266	-	-	5'GATT/G
PA-1	RAM29	HIND3	14	A71	Xq26.3	2	intron 2 of 4	8265	Z96074.1	sense	5'TCTT/A
PA-1	RAM221	HIND3	11	A56	20q13.12	1	Intron 1 of 13	8266	EYA2	sense	5'TTTT/A
PA-1	RAM272	HIND3	17	A25	16p11.1	2	intergenic	8265	-	-	5'TTTT/G
PA-1	RAM286	HIND3	12	A131	1q32.2	2	intergenic	8265	-	-	5'TTTT/C

Table 3.1: Ovarian teratocarcinoma cell lines. In this table is information regarding 5 clonally derived human ovarian teratocarcinoma cell lines. Two of the cell lines are present within introns of a gene, whereas the other 3 are present within intergenic regions.

	BS Primers	sequence 5' to 3'	Length	Product size	No. CpGs
1	p27 5' flank out-F	TtTagaattgagagTtggaaggaata	26	OUTER-292	8
*	p27 5'L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA	29		
	p27 5'flank in-F	TTagaattgagagTtggaaggaat	24	IN-233	
**	p27 5'L1 in-R	ATACCTCAAATAAAAATACAAAAATCACC	29		
2	p29 5' flank out-F	TtgagtTtgtgggTaagagagtTtg	25	OUTER-210	
*	P29 5' L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA	29		8
	P29 5' flank in-F	agagagtTtgTtggaggggaaag	23	IN-157	
**	p29 5' L1 in-R	ACTATCTAACACTCCCTAATAAAATAAACC	29		
3	p221 5' flank out-F	gaaaagtaaataaattaaaattttgagaagag	28	OUTER-355	
*	P221 5' L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA	29		
	P221 5' flank in-F	tgttgagtttaagtaatgggtgtatagaag	30	IN-315	9
**	P221 5' L1 in- R	ATACCTCAAATAAAAATACAAAAATCACC	29		
4	p272 5' flank out-F	gtaaaatTtatgaagggaTattttgg	27	OUTER-221	
*	P272 5' L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA	29		7
	p272 5' flank in-F	aaatTtatgaagggaTattttggag	25	IN-163	
**	p272 5' L1 in-R	ACTATCTAACACTCCCTAATAAAATAAACC	29		
5	p286 5' flank out-F	gtagttgttataggaggaatgtttgg	26	OUTER-346	
*	P286 5' L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA	29		7
	p286 5'flank in-F	gaggaatgtttggttaaaaataaaaag	27	IN-295	
**	p286 5'L1 in -R	ACTATCTAACACTCCCTAATAAAATAAACC	30		

Figure 3.3. BS Primers Schematic. Primers designed to amplify 5' upstream flanking DNA of L1 insertion. The outermost and innermost of the primers can be used to amplify the empty site.

*=primer sequence is the same

**=primer sequence is the same

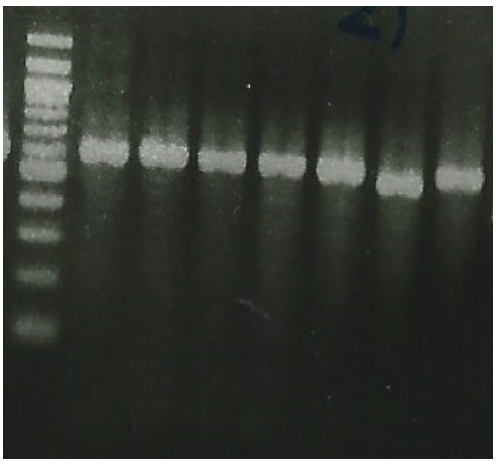
Preliminary Results:

We obtained preliminary results from DNA from subline RAM221. RAM 221 had an L1 insertion introduced into intron 1 of the EYA2 gene. We were able to find a primer combination that would amplify and clone the upstream DNA region flanking the L1 insertion. This amplicon had a total of 8 CpG sites, with 2 CpGs present within the flanking DNA itself (Fig 3.3 A). We were also able to successfully amplify and clone the empty site for RAM 221 DNA. The amplicon had a total of 5 CpG sites. Of the 5 CpG sites, 2 of them were the exact same sites amplified in the upstream DNA amplicon (Fig. 3.3 B). The other 3 CpG sites were from the region downstream of the L1 insertion. We were unable to obtain the actual methylation data from this work since COVID-19 forced the shutdown the sequencing facilities.

A.

Subline: RAM 221

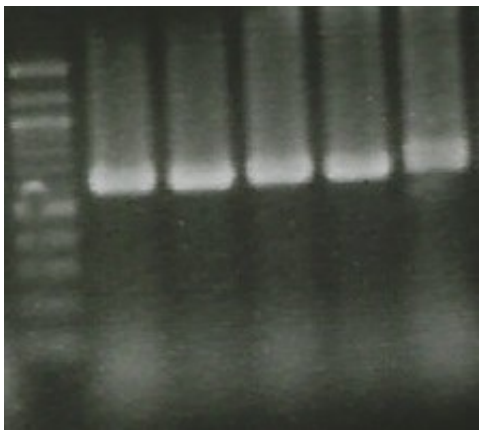
Region of Interest	*BS Primers*	*Sequence 5′ to 3′*	*Product size (bp)*	*Total CpGs*	*CpGs in flank*
5′ DNA FLANK & 5′ L1 END	p221 5' flank out-F	gaaaagtaaataaattaaaattttgagaagag	410	10	3
	P221 5′ L1 out-R	CCCACTATCTAACACTCCCTAATAAAATA			
	P221 5′ flank in-F	tgttgagtttaagtaatgggtgtatagaag	319	8	2
	P221 5′ L1 in- R	ATACCTCAAATAAAAATACAAAAATCACC			



tgtggtggagagaagagattttaagtttgaatttgttttaaatattttagtaaagagaaaagtaaataaattaaaattttgagaagagattatttggaaaa
atcgtgaatattgttgagtttaagtaatgggtgtatagaagtttattttattattttgtttgttttgtatatgttggaaatttttttattaaaaaaatttttagtttt
attatgaagtcgtttgattaaaagagaaggaggaaaaaaaaaaaattttttatcgtatgaaattgtataaaatttagttaattttaaaataagaataaaaa
tattttaaaaaaatttatgaGGGGGAGGAGTTAAGATGGTCGAATAGGAATAGTTTCGGTTTATAGTTT
TTAGCGTGAGCGACGTAGAAGACGGTGATTTTTGTATTTTTATTTGAGGTATCGGGTTTATTT
TATTAGGGAGTGTTAGATAGTGGGCGTAGGTTAGTGTGTGTGCGTATCGTGCGCGAGTCGAAGTAG
GGCGAGGTATTGTTTTATTTGGGAAGCGTAAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGG
GTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTCGAATATTGCGTTTTTTAGATCGGTTTAAGAA
ACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTACGTTTACGGAATTTCGTTGATT
GTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGGGCGTTCGTTATTG
TTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGAAGTTCGATTTGAT.....AAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAaaaaaaatttatgaatcgttgtatggttttttaatttcgggtaatgatgtaattagaggagtttagttattagaaaag
attaattttaagatatttgttgttgttagtggtagatggtagaagagtggaatattgaaagtat

B.

Region of interest	*BS Primers*	*Sequence 5′ to 3′*	*Product size (bp)*	*CpGs in flank*
EMPTY SITE (NO L1 INSERTION)	p221 5' flank out-F	gaaaagtaaataaattaaaattttgagaagag	379	5
	P221 5′ flank in-F	tgttgagtttaagtaatgggtgtatagaag	323	3
	P221 3′ flank-R	ccatctaccactaacaacaacaaatatc	-	-



tgtggtggagagaagagattttaagtttgaatttgttttaaatattttagtaaagagaaaagtaaataaattaaaattttgagaagagattatttggaaaa
atcgtgaatattgttgagtttaagtaatgggtgtatagaagtttattttattattttgtttgttttgtatatgttggaaatttttttattaaaaaaatttttagtttt
attatgaagtcgtttgattaaaagagaaggaggaaaaaaaaaaaattttttatcgtatgaaattgtataaaatttagttaattttaaaataagaataaaaa
tattttaaaaaaatttatgaGGGGGAGGAGTTAAGATGGTCGAATAGGAATAGTTTCGGTTTATAGTTT
TTAGCGTGAGCGACGTAGAAGACGGTGATTTTTGTATTTTTATTTGAGGTATCGGGTTTATTT
TATTAGGGAGTGTTAGATAGTGGGCGTAGGTTAGTGTGTGTGCGTATCGTGCGCGAGTCGAAGTAG
GGCGAGGTATTGTTTTATTTGGGAAGCGTAAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGG
GTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTCGAATATTGCGTTTTTTAGATCGGTTTAAGAA
ACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTACGTTTACGGAATTTCGTTGATT
GTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGGGCGTTCGTTATTG
TTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGAAGTTCGATTTGAT.....AAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAaaaaaaatttatgaatcgttgtatggttttttaatttcgggtaatgatgtaattagaggagtttagttattagaaaagattaatt
ttaagatatttgttgttgttagtggtagatggtagaagagtggaatattgaaagtat

Figure 3.4 Subline RAM221**.** The L1 for RAM221 is inserted into intron 1 of the EYA2 gene located at 20q13.12.

A. RAM221 DNA was bisulfite converted (see right). We successfully amplified region of the 5’ DNA flank and 5’ end of the L1 using nested PCR protocol (primers shown above). The final PCR product was ~319 bp and was cloned into a TOPO-TA vector (pCR2.1, Invitrogen) and run on a gel.

B. The empty site was also amplified using nested PCR (primers shown above). The final PCR product was ~323 bp and was cloned into a TOPO-TA vector (pCR2.1, Invitrogen) and run on a gel.

Key: lowercase is the flanking genomic DNA, uppercase is the L1 sequence, underlined are the target site duplications (TSDs), in red are the CG sites

This project began in mid-February when we first obtained the samples from our colleague Dr. Jose Luis Garcia-Perez of the University of Edinburgh. We had already designed several primers to amplify the upstream DNA flanks and the empty sites in all 5 sublines (Fig. 3.2). We wanted to successfully amplify these regions before attempting to also amplify the downstream DNA flanks in these sublines (since the presence of the poly A tail on the 3' L1 end presents a greater challenge for PCR extension). Initially, many of our primer combinations were unsuccessful, so we had to redesign new primers that would work on our bisulfite-treated DNA samples. We were starting to obtain favorable results before COVID-19 halted our progress.

The successful completion of this project would have required methylation data from CpG sites both upstream and downstream of where the L1 insertion is in all the sublines. This data must include information about the methylation status of the same flanking CpG sites in the chromosome with the insertion and the chromosome without it (the empty site). We had planned to obtain this data by amplifying the flanking CpG sites using the primer combinations listed in Figure 3.2. We were successful in amplifying the empty site and 5'flanking DNA of subline RAM221 before COVID-19 halted our progress. With the entirety of our data we would have been able to do bisulfite conversion methylation analysis by using the Quantification tool for Methylation Analysis (QUMA). QUMA is a commonly used program that uses Fisher's Exact Test statistics to determine whether any differences in methylation levels are statistically significant [79]. QUMA would have allowed us to determine whether the presence of the L1 insertion led to statistically significant changes in the average methylation levels observed in each flanking CpG site (both upstream and downstream from the insertion).

We were expecting the average methylation levels of the flanking CpG sites to be significantly elevated in the presence of the L1 insertion relative to the empty site. We expected

for this to be true for CpG sites located both upstream and downstream of the insertion. We expected for this to be the case for all 5 sublines, regardless of whether the L1 insertion was present in an intergenic or intronic region. This result would have supported our hypothesis that new L1 insertions can affect the methylation status of flanking DNA and for those L1s inserted in genes, to investigate effects on gene expression. This preliminary data has laid the groundwork for this and the above studies to be continued at a later date following the reopening of the university.

References:

1. McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci. 1950;36(6):344–355. doi: 10.1073/pnas.36.6.344.
2. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011;7:e1002384. doi: 10.1371/journal.pgen.1002384.
3. Goodier JL, Kazazian HH. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. 2008;135(1):23–35. doi: 10.1016/j.cell.2008.09.022.
4. Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res. 2008;18:717–28
5. Kazazian, H. H. (2011). Mobile Dna: finding treasure in junk. Upper Saddle River, N.J: Financial Times/Prentice Hall.
6. Goodier, John L. "Restricting retrotransposons: a review." Mobile DNA vol. 7 16. 11 Aug. 2016, doi:10.1186/s13100-016-0070-z
7. Cordaux R., Batzer M.A. 2009, The impact of retrotransposons on human genome evolution, Nat. Rev. Genet., 10, 691–703.
8. Lander E.S., Linton L.M., Birren B., et al. 2001, Initial sequencing and analysis of the human genome, Nature, 409, 860–921.
9. Wanxiangfu Tang, Seyoung Mun, Aditya Joshi, Kyudong Han, Ping Liang, Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase, DNA Research, Volume 25, Issue 5, October 2018, Pages 521–533
10. . E. M. Ostertag, H. H. Kazazian Jr., Annu. Rev. Genet. 35, 501 (2001)
11. Ewing, A. D., & Kazazian, H. H., Jr (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome research, 20(9), 1262–1270. https://doi.org/10.1101/gr.106419.110
12. Wessler, Susan R. "Transposable elements and the evolution of eukaryotic genomes." Proceedings of the National Academy of Sciences of the United States of America vol. 103,47 (2006): 17600-1. doi:10.1073/pnas.0607612103
13. Hancks DC, Kazazian Jr HH. Active human retrotransposons: variation and disease. Curr Opin Genet Dev. 2012;22:191–203

14. Craig NL. Unity in transposition reactions. Science. 1995;270(5234):253–4.
15. Craig NL. Target site selection in transposition. Annu Rev Biochem. 1997;66:437–74.
16. Kazazian H.H. 2004. Mobile elements: Drivers of genome evolution. Science. 303, 1626–1632.
17. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci. 2016
18. Xiong, Y. & Eickbush, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362 (1990).
19. Boeke, J. D., Garfinkel, D. J., Styles, C. A., & Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. Cell, 40(3), 491-500. https://doi.org/10.1016/0092-8674(85)90197-7
20. Lueders KK, Kuff EL. Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell. 1977;12(4):963–72.
21. Havecker, E.R., Gao, X. & Voytas, D.F. The diversity of LTR retrotransposons. Genome Biol 5, 225 (2004). https://doi.org/10.1186/gb-2004-5-6-225
22. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921
23. Hancks, D.C., Kazazian, H.H. Roles for retrotransposon insertions in human disease. Mobile DNA 7, 9 (2016). https://doi.org/10.1186/s13100-016-0065-9
24. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian Jr HH. High Frequency Retrotransposition in Cultured Mammalian Cells. Cell. 1996; 87:917–27
25. Viollet, S., Monot, C., & Cristofari, G. (2014). L1 retrotransposition: The snap-velcro model and its consequences. Mobile genetic elements, 4(1), e28907. https://doi.org/10.4161/mge.28907
26. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. Cell. 1993;72:595–605

27. Ostertag, E. M., & Kazazian, H. H., Jr (2001). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome research, 11(12), 2059–2065. https://doi.org/10.1101/gr.205701
28. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 2003;35:41–8.
29. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet. 2011;20:3386–400
30. Kazazian, H. H., Jr, & Moran, J. V. (2017). Mobile DNA in Health and Disease. The New England journal of medicine, 377(4), 361–370. https://doi.org/10.1056/NEJMra1510092
31. Created with BioRender.com
32. Batzer M.A., Deininger P.L.. Alu repeats and human genomic diversity, Nat. Rev. Genet., 2002, vol. 3 (pg. 370-379)
33. Wallace N., Wagstaff B.J., Deininger P.L., Roy-Engel A.M.. LINE-1 ORF1 protein enhances Alu SINE retrotransposition, Gene, 2008, vol. 419 (pg. 1-6)
34. Kazazian, H., Wong, C., Youssoufian, H. et al. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332, 164–166 (1988). https://doi.org/10.1038/332164a0
35. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y Cancer Res. 1992 Feb 1; 52(3):643-5.
36. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 2012;40(4):1666–83.
37. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. Jr. Exon shuffling by L1 retrotransposition. Science 283, 1530–1534 (1999)
38. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. SVA elements are nonautonomous retrotransposons that cause disease in humans. The American Journal of Human Genetics. 2003;73(6):1444-5
39. Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, Russell DW. 1985. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227:140–46
40. Bangsbøll S, Qvist I, Lebech PE, Lewinsky M. Testicular feminization syndrome and associated gonadal tumors in Denmark. Acta Obstet Gynecol Scand. 1992;71:63–66.
41. Rafael Loch Batista, Katsumi Yamaguchi, Andresa di Santi Rodrigues, Mirian Yumie Nishi, John L Goodier, Luciani Renata Carvalho, Sorahia Domenice, Elaine M F Costa, Haig H Kazazian, Jr, Berenice Bilharinho Mendonca, Mobile DNA in Endocrinology: LINE-1 Retrotransposon Causing Partial Androgen Insensitivity Syndrome, The Journal of Clinical

Endocrinology & Metabolism, Volume 104, Issue 12, December 2019, Pages 6385–6390, https://doi.org/10.1210/jc.2019-00144
42. Hornig, N. C., Ukat, M., Schweikert, H. U., Hiort, O., Werner, R., Drop, S. L., Cools, M., Hughes, I. A., Audi, L., Ahmed, S. F., Demiri, J., Rodens, P., Worch, L., Wehner, G., Kulle, A. E., Dunstheimer, D., Müller-Roßberg, E., Reinehr, T., Hadidi, A. T., Eckstein, A. K., … Holterhus, P. M. (2016). Identification of an AR Mutation-Negative Class of Androgen Insensitivity by Determining Endogenous AR Activity. The Journal of clinical endocrinology and metabolism, 101(11), 4468–4477. https://doi.org/10.1210/jc.2016-1990
43. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, Moran JV Nature. 2010 Aug 5; 466(7307):769-73.
44. Bragg et al., 2017 D.C. Bragg, K. Mangkalaphiban, C.A. Vaine, N.J. Kulkarni, D. Shin, R. Yadav, J. Dhakal, M.L. Ton, A. Cheng, C.T. Russo, et al. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1 Proc Natl Acad Sci U S A. (2017)
45. Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. J Virol. 2014;88(17):9529–37.
46. Naveira H, Bello X, Abal-Fabeiro JL, Maside X. Evidence for the persistence of an active endogenous retrovirus (ERVE) in humans. Genetica. 2014;142(5):451–60.
47. Martin S. L. (2010). Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA biology, 7(6), 706–711. https://doi.org/10.4161/rna.7.6.13766
48. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Löwer, J., Strätling, W. H., Löwer, R., & Schumann, G. G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic acids research*, *40*(4), 1666–1683. https://doi.org/10.1093/nar/gkr863
49. Pickeral OK, Makalowski W, Boguski MS, Boeke JD (2000) Frequent genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 10:411–41510.1101/gr.10.4.411
50. Moran JV, DeBerardinis RJ, Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. Science 283:1530–153410.1126/science.283.5407.1530
51. Goodier JL, Ostertag EM, Kazazian HH Jr (2000) Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9:653–65710.1093/hmg/9.4.653
52. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16: 276-277.
53. Xie, H., Wang, M., Bonaldo, Mde. F., Smith, C., Rajaram, V., Goldman, S., Tomita, T., and Soares,M.B. (2009) High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. Nucl. Acids Res. 37: 4331-4340
54. Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13: 335–339.
55. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: a hominid-specific retroposon family. J Mol Biol. 2005;354(4):994-1007. doi: 10.1016/j.jmb.2005.09.085

56. Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol. 2010;20(4):211-21.
57. De Cecco M, Criscione SW, Peckham EJ, Hillenmeyer S, Hamm EA, Manivannan J, Peterson AL, Kreiling JA, Neretti N, Sedivy JM. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell. 2013;12(2):247-56.
58. Arnaud, P., Goubely, C., Pelissier, T., and Deragon, J.M. (2000) SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol. Cell. Biol. 20: 3434-3441.
59. Saze, H. and Kakutani, T. (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. EMBO J. 26: 3641–3652.
60. Graff, J.R., Herman, J.G., Myöhänen, S., Baylin, S.B., and Vertino, P.M. (1997). Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. J. Biol. Chem. 272: 22322-22329.
61. Yates, P.A., Burman, R.W., Mummaneni, P., Krussel, S., and Turker, M.S. (1999) J. Biol. Chem. 274: 36357–36361
62. Fukuda K, Inoguchi Y, Ichiyanagi K, Ichiyanagi T, Go Y, Nagano M, Yanagawa Y, Takaesu N, Ohkawa Y, Imai H, Sasaki H. Evolution of the sperm methylome of primates is associated with retrotransposon insertions and genome instability. Hum Mol Genet. 2017;26(18):3508-19.
63. Grandi FC, Rosser JM, Newkirk SJ, Yin J, Jiang X, Xing Z, Whitmore L, Bashir S, Ivics Z, Izsvak Z, Ye P, Yu YE, An W. Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. Genome research. 2015;25(8):1135-46
64. Levin, H., Moran, J. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12, 615–627 (2011). https://doi.org/10.1038/nrg3030
65. Lee LV, et al. (2011) The unique phenomenology of sex-linked dystonia parkinsonism (XDP, DYT3, “Lubag”) Int J Neurosci 121:3–11.
66. Lee LV, Kupke KG, Caballar-Gonzaga F, Hebron-Ortiz M, Müller U (1991) The phenotype of the X-linked dystonia-parkinsonism syndrome. An assessment of 42 cases in the Philippines. Medicine (Baltimore) 70:179–187.
67. Fahn S, Moskowitz CB. X-linked recessive dystonia and parkinsonism in Filipino males. Ann Neurol 1988;24:179.
68. Anandapadamanaban M, et al. (2013) High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. Nat Struct Mol Biol 20: 1008–1014.
69. Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105–1
70. Domingo A, et al. (2015) New insights into the genetics of X-linked dystoniaparkinsonism (XDP, DYT3). Eur J Hum Genet 23:1334–1340
71. Wilhelmsen KC, et al. (1998) Molecular genetic analysis of Lubag. Adv Neurol 78:341–348
72. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 2009;19(11):1983-91.

73. Bragg, D. C., Mangkalaphiban, K., Vaine, C. A., Kulkarni, N. J., Shin, D., Yadav, R., Dhakal, J., Ton, M. L., Cheng, A., Russo, C. T., Ang, M., Acuña, P., Go, C., Franceour, T. N., Multhaupt-Buell, T., Ito, N., Müller, U., Hendriks, W. T., Breakefield, X. O., Sharma, N., … Ozelius, L. J. (2017). Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. Proceedings of the National Academy of Sciences of the United States of America, 114(51), E11020–E11028. https://doi.org/10.1073/pnas.1712526114
74. Westenberger A, Reyes CJ, Saranza G, Dobricic V, Hanssen H, Domingo A, Laabs BH, Schaake S, Pozojevic J, Rakovic A, Grutz K, Begemann K, Walter U, Dressler D, Bauer P, Rolfs A, Munchau A, Kaiser FJ, Ozelius LJ, Jamora RD, Rosales RL, Diesta CCE, Lohmann K, Konig IR, Bruggemann N, Klein C. A hexanucleotide repeat modifies expressivity of X-linked dystonia-parkinsonism. Ann Neurol. 2019. Epub
75. Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol. Cell. Biol. 25 7780-7795.
76. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115. https://doi.org/10.1093/nar/gks596
77. Li LC and Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics. 2002 Nov;18(11):1427-31. PMID: 12424112.
78. Kint, S., De Spiegelaere, W., De Kesel, J., Vandekerckhove, L., & Van Criekinge, W. (2018). Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. PloS one, 13(6), e0199091. https://doi.org/10.1371/journal.pone.0199091
79. Kumaki Y, Oda M, Okano M. QUMA: quantification tool for methylation analysis. Nucleic Acids Res. 2008;36(Web Server issue):W170–W175. doi: 10.1093/nar/gkn294.
80. Kint, S., De Spiegelaere, W., De Kesel, J., Vandekerckhove, L., & Van Criekinge, W. (2018). Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. PloS one, 13(6), e0199091. https://doi.org/10.1371/journal.pone.0199091
81. Shvetsova, E., Sofronova, A., Monajemi, R. et al. Skewed X-inactivation is common in the general female population. Eur J Hum Genet 27, 455–465 (2019). https://doi.org/10.1038/s41431-018-0291-3

Clement Ojo
| 401-660-3098 |
clementojo1193@gmail.com

Education

Johns Hopkins University-Krieger School of Arts and Sciences Baltimore, MD

- Bachelor of Science: Molecular and Cellular Biology, Psychology Minor (Honors) Class of 2019
- Pursuing MS degree in Molecular Biology

Teaching Experiences

- Teaching Assistant for General Biology Laboratory I -Fall 2019
- Teaching Assistant for Cell Biology Lab- Spring 2020
- Cellist teacher for middle school aged children- Fall 2012-2017

Honors/Awards

- Dean's List Fall 2016-Spring 2020
- National Society of Collegiate Scholars-invite-Fall 2017
- Omicron Kappa Honor Society-invitation-Spring 2019
- TriBeta Honor Society – Member- Spring 2019-Present

Research Experiences

- Undergraduate Research
  - Kazazian Lab – Studied L1 elements in gastrointestinal cancers (Fall 2017-Spring 2019)
  - Skolasky Lab- Clinical research on shared decision making and treatment expectations in spine surgery (Fall 2018-Spring 2019)
- Graduate Research
  - Kazazian Lab- Studied the effects of non-LTR retrotransposon insertion on the epigenetic profile of the integration site.

Other Experiences

- **Beloved College Community Fellowship-** Fall 2016- 2019
  - Member and Service Leader, Community Service involvement
- **Warwick Symphony Orchestra-** Fall 2012-2017
  - First Chair Cellist (professional orchestra member)
- **Musicare-** Fall 2018-2019
  - Volunteer Cellist. I played at hospitals & senior centers
- **Shaking Hands-** Fall 2017-2020
  - Vice President/Co-founder. Tutor for children.

- **African Student Association-** Fall 2016-2019
    - Active member and participant in culture events
- **Black Student Union** Fall 2016- 2019
    - Active Member. Knowledgeable about issues affecting black students at Hopkins
- **Igbara-Oke Citizens in the Americans, Inc –** Fall 2016-Now
    - Organizes projects of economic and social benefit to people of Igbara-Oke village and surrounding Nigerian communities. I help plan some activities, fundraisers, and participate in events.
- **Johns Hopkins Underrepresented in Medical Professions-** Fall 2016-2019
    - Member
- **Supporting Hospitals Abroad with Resources+Equipment-** Fall 2016-2018
    - We sort unused/unopen hospital materials and package them abroad to be used in other countries.
- **Wingate Assisted Living- Memory Care Unit –** Member Spring-Summer 2018
    - I worked in Wingate's Memory Care unit making beds, serving meals, and interacting with the residents. I worked with licensed nurse practitioners, physical therapists and CNAs.
- **Rhode Island for Community and Justice (RICJ)** Summer 2018
    - Trained to and have led several difficult conversations around topics such as race, gender, drug abuse, in the youth of the Rhode Island Community. I have organized workshops.
- **Infectious Disease Center-Mariam Hospital -**Volunteer/Shadowing Summer 2019
    - I shadowed Dr. Timothy Flanagan, an infectious disease specialist at Miriam Hospital. I interacted with patients and assisted them with basic needs.
- **STD/HIV & Lyme Disease Clinics –** Summer 2019
    - I shadowed Dr. Michaela Maynard and observed the interactions between the patients and the healthcare provider.
- **Johns Hopkins Department of Orthopedic Surgery-**Shadowing Fall 2017
    - Shadowed Dr. Khaled Kebaish, the division chief of Orthopedic Spine Surgery at Johns Hopkins Hospital.